

Basalkursus, opgaver vedr. simpel lineær regression

Opgave 1: Resting Metabolic Rate

I filen (`rmr.txt`) findes sammenhørende værdier af kropsvægt (**bw, i kg**) og hvilende stofskifte (**rmr, kcal pr. døgn**) for 44 raske kvinder (Altman, 1991 og Owen et.al., 1986). Første linie indeholder variabelnavnene.

Vi ønsker at undersøge, hvordan kropsvægten påvirker stofskiftet:

1. Lav en tegning af de sammenhørende værdier af kropsvægt og stofskifte. Overvej i denne forbindelse, hvad der bør være X- hhv. Y-akse.
2. Opstil en rimelig model (evt. ved først at transformere data på passende vis, hvis du synes, modelforudsætningerne halter), og estimer parametrene i denne, med tilhørende spredninger (spredninger på parameterestimater = standard errors).
3. Hvad er estimatet for middelværdien af hvilestofskiftet for kvinder på 70 kg?
Hvis det tilsvarende middel hvilestofskifte for mænd (på 70 kg) vides at være skønnet til 1406 kcal/døgn (med CI på 1360-1452), kan vi så konkludere, at der er forskel på hvilestofskiftet hos mænd og kvinder på 70 kg?
(Dette spørgsmål kan evt. besvares løseligt ud fra en tegning med indlagte konfidensgrænser).
4. I ambulatoriet ser vi en kvinde, der vejer 80 kg.
 - Hvad er dit bedste gæt på hendes hvilestofskifte (i mangel af yderligere information)? Og indenfor hvilke grænser vil du med 95% sikkerhed tippe hendes hvilestofskifte at ligge, forudsat at hun er rask.
 - Hvis hun viser sig at have et hvilestofskifte på 950 kcal/døgn, hvilken diagnose ville du så stille, og hvorfor?
 - Er det muligt (med 95% sikkerhed) at forudsige en enkelt kvindes stofskifte udfra kropsvægten med en nøjagtighed på ± 250 kcal/døgn ?

(Til dette spørgsmål kan man ligeledes benytte et plot, blot med prediktionsgrænser i stedet for konfidensgrænser).

5. Vurder om modellens forudsætninger kan siges at være opfyldt. Suppler evt. med teoretiske overvejelser omkring stofskifte.
Til dette spørgsmål (samt til det følgende) kan man evt. danne et nyt datasæt ved at benytte en **output**-sætning med angivelse af, hvilke ekstra informationer, man ønsker gemt. Alternativt kan man benytte faciliteterne i **ods**-systemet.
6. Overvej, om der muligvis er enkelte observationer, der har en speciel kraftig indflydelse på estimationen. Hvad kan/bør man gøre med sådanne?

Hvis der bliver mere tid, kan I forlyste jer med:

Opgave 2: Definition af BMI

Vi skal se på rimeligheden i definitionen af body mass index

$$\text{BMI} = \frac{\text{vægt i kg}}{(\text{højde i m})^2}$$

og skal hertil benytte Sundby-data.

1. Transformer ovenstående teoretiske relation (altså selve formelen) med logaritmen, dvs. find et teoretisk udtryk for logaritmen til BMI.
Hvis alle havde samme BMI, hvilken lineær relation ville vi så forvente at finde mellem de logaritmerede værdier af vægt og højde?
2. Hent nu **sundby** ind, enten fra **sasuser**, hvis den findes der fra tidligere, eller fra filen på hjemmesiden.
3. Tegn et scatterplot af logaritmen til vægt (vægt er **v75**) overfor logaritmen til højde (højde er **v76**), for hvert køn for sig, og vurder forudsætningerne for at foretage en lineær regression.
4. Fit en lineær relation, med $\log(\text{vægt})$ som respons og $\log(\text{højde})$ som kovariat, for et af kønnene (eller hvert køn for sig).
Hvordan passer resultatet med definitionen af bmi?
5. Hvordan kunne man forklare en evt. afvigelse fra det forventede?